Illuminating protein space with a programmable generative model

Year: 2022 | Citations: 465 | Authors: John Ingraham, Max Baranov, Zak Costello, Vincent Frappier, Ahmed Ismail

---

Abstract

Three billion years of evolution has produced a tremendous diversity of protein molecules^ 1 , but the full potential of proteins is likely to be much greater. Accessing this potential has been challenging for both computation and experiments because the space of possible protein molecules is much larger than the space of those likely to have functions. Here we introduce Chroma, a generative model for proteins and protein complexes that can directly sample novel protein structures and sequences, and that can be conditioned to steer the generative process towards desired properties and functions. To enable this, we introduce a diffusion process that respects the conformational statistics of polymer ensembles, an efficient neural architecture for molecular systems that enables long-range reasoning with sub-quadratic scaling, layers for efficiently synthesizing three-dimensional structures of proteins from predicted inter-residue geometries and a general low-temperature sampling algorithm for diffusion models. Chroma achieves protein design as Bayesian inference under external constraints, which can involve symmetries, substructure, shape, semantics and even natural-language prompts. The experimental characterization of 310 proteins shows that sampling from Chroma results in proteins that are highly expressed, fold and have favourable biophysical properties. The crystal structures of two designed proteins exhibit atomistic agreement with Chroma samples (a backbone root-mean-square deviation of around 1.0 Å). With this unified approach to protein design, we hope to accelerate the programming of protein matter to benefit human health, materials science and synthetic biology. Evolution has produced a range of diverse proteins, and now a generative model called Chroma can expand that set by allowing the user to design new proteins and protein complexes with desired properties and functions.